

Langues et Gènes en Asie orientale

SAGART Laurent & POLONI Estella

Action “ Origine de l’Homme, du Langage et des Langues ”

A. FICHE ADMINISTRATIVE

Titre du projet :

Langues et Gènes en Asie orientale

Mots-clés :

.....
Linguistique; génétique; Asie orientale; Gm, Rhésus, migrations; histoire du peuplement.
.....

Résumé du projet (10 lignes maximum) :

On examinera l'histoire du peuplement de l'Asie orientale en confrontant les principales hypothèses de classification des langues d'Asie orientale, représentées par des arbres datés, avec les affinités génétiques des populations, étudiées sur la base des systèmes Gm et Rhésus. Les données génétiques seront extraites d'une base de données rassemblée par l'une d'entre nous (A. S-M). Après avoir conduit une analyse globale des affinités génétiques entre populations, on estimera la corrélation entre distances linguistiques et distances génétiques, indépendamment de la géographie; on testera la perméabilité des frontières linguistiques au flux génique, ainsi qu'aux emprunts linguistiques; l'étude des faits de contact sous deux angles, génétique et linguistique, permettra d'affiner le scénario historique. Dans un second temps, la même démarche sera appliquée à d'autres systèmes génétiques. Un terrain est à prévoir afin de compléter les données.

1. Responsable scientifique du projet

Nom : SAGART..... Prénom ...Laurent.....

Grade : Directeur de recherche.....

Discipline du responsable scientifique: linguistique.....

Établissement de rattachement : CNRS.....

Adresse professionnelle : N° 54, Bd Raspail
Code postal : 75270 Commune : Paris cédex 06

Tél +33-1-49 54 24 18 Fax +33-1-49 54 26 71

E-Mail : sagart@ehess.fr

1. Co-responsable scientifique du projet

Nom POLONI Prénom Estella

Grade Maitre-Assistante

Discipline du responsable scientifique Génétique des populations

Établissement de rattachement

Laboratoire de Génétique et Biométrie
Département d'Anthropologie et Ecologie de l'Université de Genève

Adresse professionnelle :

Département d'Anthropologie et Ecologie
12, rue Gustave-Revilliod
C.P. 511, 1211 Genève 24, Suisse

Tél +41 22 702 69 77 Fax +41 22 300 03 51

E-Mail estella.poloni@anthro.unige.ch

2. Laboratoire ou organisme de rattachement de l'équipe de recherche

Intitulé Centre de Recherches Linguistiques sur l'Asie Orientale (CRLAO)

Type de formation (*cocher la case utile*)

- Unités CNRS : unité associée ou mixte du CNRS

Préciser le code unité *UMR 8563*
Préciser la délégation régionale : Paris A

Nom du directeur de l'organisme : R. Djamouri, A. Peyraube

Adresse : N° 54, Bd Raspail
Code postal 75270 Commune Paris Cédex 06

Tél +33-1-49 54 24 18 Fax +33-1-49 54 26 71

3. Autre(s) laboratoires ou organisme(s) partenaires

3.1

Intitulé : Laboratoire de Génétique et Biométrie Département d'Anthropologie et Ecologie de l'Université de Genève

Discipline(s) couverte(s) par l'équipe : Génétique des Populations

Nom du directeur de l'équipe : Prof. A. Langaney

Membre(s) partenaire(s) :

1. Nom : Poloni
Prénom : Estella
Grade : Maître-assistante

- E-Mail : estella.poloni@anthro.unige.ch
2. Nom : Sanchez-Mazas
Prénom : Alicia
Grade : Maître d'enseignement et de recherche
E-Mail : alicia.sanchez-mazas@anthro.unige.ch

Adresse : C.P. 511, 1211 Genève 24 / Suisse

Tél +41(22)7026977

Fax +41(22)3000351

3.2

Intitulé : Dipartimento di Biologia, Università di Ferrara

Discipline(s) couverte(s) par l'équipe : Génétique des Populations

Nom du directeur de l'équipe : Prof. G. Barbujani

Membre partenaire :

Nom : Dupanloup de Ceuninck

Prénom : Isabelle

Grade : Post-doctorante

E-Mail : isabelle.dupanloup@anthro.unige.ch

Adresse : Via L. Borsari 46, 44100 Ferrara, Italie

Tél. +39 0532 291743

Fax +39 0532 249761

Date : 9 Juin 2000

Signature du Responsable Scientifique :

B. PROJET SCIENTIFIQUE

B.1. Introduction

Depuis la fin des années quatre-vingt, les généticiens des populations ont pu observer que les affinités génétiques entre populations humaines actuelles tendent à correspondre aux relations de ces populations, inférées par la linguistique historique (e.g. Cavalli-Sforza et al. 1994), ceci au moins dans les cas où les relations génétiques font l'objet d'un consensus de la part des linguistes. Cette observation suggère, d'une manière générale, un parallélisme entre les processus évolutifs de différenciation génétique et de différenciation linguistique des populations humaines. La confrontation des résultats et des hypothèses tirés de ces deux disciplines — génétique des populations et linguistique historique — permet donc de suggérer des scénarios relatifs à l'histoire du peuplement humain. Nous présentons ici un projet de recherche faisant appel à ces deux disciplines, dans le cadre de l'histoire du peuplement humain de l'Asie orientale.

B.2. Histoire du peuplement de l'Asie orientale

Diverses disciplines permettent d'aborder l'histoire du peuplement humain. Nous nous intéresserons ici à la linguistique diachronique et la génétique des populations, deux disciplines

pour lesquelles des spécialistes se sont réunis dans le présent projet. Ci-dessous nous présentons un bref survol des hypothèses d'origine, d'apparentement et/ou de différenciation des langues et des profils génétiques des populations de l'Asie orientale. Les hypothèses tirées d'autres disciplines, telles que l'archéologie et la paléanthropologie ne sont pas abordées ici, mais seront utilisées dans la phase d'interprétation des résultats obtenus dans le cadre du présent projet.

B.2.1 Hypothèses sur les apparentements linguistiques en Asie orientale

La phylogénie ancienne des langues d'Asie Orientale est très controversée. Plusieurs hypothèses sont actuellement “ sur le marché ”. Ce sont :

(1) L'austrique (Schmidt 1906) : austronésien + austroasiatique. Cette hypothèse a été récemment soutenue par Reid (1994). Les preuves en sont surtout morphologiques : il existe peu de lexique commun.

(2) L'austro-tai (Benedict 1942, 1975) : austronésien + kadai + miao-yao. Le berceau serait en Chine du sud-est.

(3) L'austro-japonais : austro-tai + japonais (Benedict; divers auteurs japonais, comme Murayama, proposent une version limitée à l'austronésien et au japonais, excluant le kadai). D'autres auteurs (Vovin 1994) considèrent que le japonais est apparenté au coréen, et, au-delà, à l'altaïque.

(4) Le macro-austrique : austronésien + kadai + austroasiatique (par ex. Ruhlen, sous le nom d' "austrique")

(5) Le sino-austronésien : chinois + tibéto-birman + austronésien (Sagart 1994) ; le thai est considéré comme une langue austronésienne tardive, retournée sur le continent, et ayant été en partie relexifiée. La langue ancestrale est placée dans la région de la domestication du millet vers -7000.

(6) Le sino-caucasien : sino-tibétain + ket (Yeniseï) + nord caucasien (Starostin 1984). Berceau occidental ; la langue ancestrale aurait été parlée jusqu'à -8000 approximativement.

(7) Le sino-indo-européen : sino-tibétain + indo-européen (Pulleyblank 1995).

(8) Starosta (2000 ; non publié) propose un proto-est-asiatique qui associe le sino-austronésien de Sagart et l'austrique de Reid comme deux branches d'un vaste ensemble est-asiatique. Berceau dans la région des deux domestications riz (fleuve bleu), millet (fleuve jaune). La dispersion de la famille serait antérieure à ces domestications.

L'une des raisons de cette divergence d'opinions est l'existence dans la région de très importants phénomènes de contact qui ont abouti au transfert de formes de tous ordres d'une langue à l'autre. Ceci tend à brouiller les faits de parenté génétique réelle. Ainsi, il est maintenant bien admis que le chinois, langue d'une culture dominante dans la région depuis quelque 3000 ans, a profondément influencé les langues en contact avec lui, en particulier celles parlées plus au sud. De même, il existe de fortes indications d'une forte influence des langues altaïques sur le chinois, plus particulièrement le chinois du nord (Hashimoto 1986). Des contacts plus localisés entre

langues tibéto-birmanes, tai, austroasiatiques et miao-yao ont également eu lieu.

B.2.2. Hypothèses sur le peuplement de l'Asie orientale tirées de la génétique

L'analyse de la différenciation génétique des populations d'Asie orientale a été abordée par l'étude de différents systèmes génétiques sérologiques (dits "classiques"). D'une manière générale, les différenciations génétiques des populations de l'Asie orientale semblent être fortement structurées selon une composante géographique, même si celle-ci n'a pas été explicitement testée. Cette composante peut être expliquée par l'existence de foyers originels de peuplement différenciés, puis ultérieurement homogénéisés par intermigrations, ou au contraire en invoquant une origine unique du peuplement de cette région mondiale. Ainsi, par exemple, Cavalli-Sforza et al. (1994) présentent une synthèse sur la base de différents systèmes sérologiques confondus, tout en insistant sur le fait que les données génétiques sont encore trop rares dans cette région mondiale. Cette synthèse met en avant une différenciation marquée entre nord-est asiatiques et sud-est asiatiques, partiellement effacée par "métissages graduels" (Cavalli-Sforza et al. 1994, p. 254).

Pour le système Gm en particulier, les analyses les plus récentes (Matsumoto 1988a,b; Zhao & Li 1989; revues dans Sanchez-Mazas 1990) mettent en évidence un gradient général de variation continue des fréquences géniques sur l'Asie orientale selon un axe nord-sud. Pour expliquer les résultats, certains chercheurs invoquent un scénario faisant appel à deux foyers différenciés de peuplement placés en Chine (l'un au nord, l'autre au sud) ou à sa périphérie. L'apparition du gradient de variation serait conséquent aux migrations Han du nord vers le sud au cours des deux derniers millénaires, en raison d'échanges génétiques entre les migrants et les populations locales. Ce scénario est contrasté par les analyses de Sanchez-Mazas (1990) qui incluent également des populations d'Asie occidentale. Les résultats révèlent cette fois deux grands axes de relations génétiques continues entre populations, de l'Asie occidentale vers l'Asie orientale, contournant l'Himalaya par le nord et par le sud, respectivement, et qui vient donc nuancer le scénario à deux foyers originels. Ce même résultat est obtenu par l'analyse des haplotypes du système Rhésus (Sanchez-Mazas 1990).

Plus récemment, des analyses basées sur des polymorphismes moléculaires ont commencé à apparaître. On peut séparer ces études en deux catégories:

- la première catégorie comprend les études génétiques centrées sur la problématique du peuplement d'autres régions mondiales, et ne s'intéressent à la variabilité génétique de l'Asie orientale que comme source d'origine pour d'autres populations (e.g. pour l'Amérique: Karafet et al. 1997, 1999; pour l'Océanie: Melton et al. 1998, Hagelberg et al. 1999; pour l'Océanie; pour l'Asie centrale et le nord de l'Europe: Zerjal et al. 1997).
- la seconde catégorie comprend les études centrées sur la problématique du peuplement de l'Asie orientale, sur la base des relations génétiques entre populations du territoire de la République Populaire de Chine (Chu et al. 1998; Su et al. 1999). Ces populations ont l'avantage de représenter pratiquement toutes les familles linguistiques d'Asie orientale, dont l'Austronésien si

l'on compte Taiwan. De cette seconde catégorie, sont issues deux hypothèses: la première supporte le scénario à deux foyers inféré d'après l'étude des polymorphismes classiques (Chu et al. 1998), alors que la seconde invoque une origine unique et ancienne du peuplement en Asie du sud-est, puis remontée vers le nord en Asie orientale (Su et al. 1999).

D'autres études, enfin, s'intéressent au peuplement d'une région spécifique de l'Asie orientale, comme le Japon (e.g. Hammer & Horai 1995) ou la Corée (e.g. Kim et al. 2000). Toutes ces études, cependant, n'ont abordé que de manière très superficielle les relations pouvant exister entre processus de différenciation linguistique et génétique.

B.3. Confrontation des deux disciplines

Comme nous venons de le voir, la relation entre différenciation génétique et linguistique en Asie orientale n'a pour l'heure pas été abordée telle quelle. D'où le but du présent projet.

B.3.1. Objectifs scientifiques

Nous nous proposons de confronter les diverses hypothèses linguistiques en présence sur les regroupements de niveau supérieur entre langues d'Asie orientale aux hypothèses de relations génétiques entre les populations parlant ces langues, inférées d'après l'étude de la variabilité génétique des populations actuelles. Notre intérêt réside dans la corrélation qui peut exister (ou non) entre les processus évolutifs de différenciation génétique et de différenciation linguistique des populations d'Asie orientale, ceci indépendamment des facteurs géographiques qui peuvent avoir influencé (en termes de contacts et de flux de gènes) les profils génétiques de ces populations. Ceci implique que nous incorporions, dans nos analyses, la composante géographique de la variabilité linguistique et génétique.

La faisabilité de notre projet dépend de manière cruciale de l'échantillonnage de populations humaines disponible, ceci afin de disposer d'une bonne représentativité des différents groupes linguistiques de cette région du monde. Pour l'heure, deux systèmes génétiques sérologiques répondent à cette condition, Gm et Rhésus, dont la distribution de la variabilité haplotypique dans les populations mondiales a été compilée en une base de données par A. Sanchez-Mazas (Sanchez-Mazas 1990; Sanchez-Mazas & Pellegrini 1990).

B.3.2. Matériel

Du point de vue linguistique, on se propose d'obtenir des principaux avocats de chacune des théories linguistiques des arbres phylogénétiques datés, exprimant les relations de parenté supposées entre les langues d'Asie orientale, dans lesquels les principaux nœuds seraient approximativement datés.

Du point de vue génétique, les profils génétiques des populations d'Asie orientale qui ont été étudiées pour les systèmes Gm et Rhésus sont stockés dans une base de données mondiales (Sanchez-Mazas 1990). Chaque entrée dans la base de données est structurée en un certain nombre de champs dont:

- une description détaillée de la population étudiée (ethnie, pays et continent, localisation précise, langue, taille de la population),
- une description détaillée de l'échantillon étudié,
- la référence bibliographique de l'étude dont sont tirées les données,
- ainsi qu'un certain nombre de rubriques permettant le classement et le tri des données.

La dernière mise-à-jour de la base de données sur les polymorphismes Gm et Rhésus comprend toutes les données publiées jusqu'en 1992. Sur cette base, un survol préliminaire des données directement utiles au projet est résumé dans la table ci-dessous. D'autres données sont également disponibles, notamment pour des régions géographiques voisines et/ou pour d'autres familles linguistiques (par exemple, l'Indo-Européen).

La première phase du projet consistera donc à compléter la base avec les données publiées depuis 1992, ainsi que des données non publiées mais qui nous seront communiquées par des spécialistes. Nous accèderons également à des données publiées en chinois, qui seront traduites par les linguistes.

Nous profiterons de cette phase de recherche bibliographique pour évaluer la disponibilité de données sur d'autres systèmes génétiques (sérologiques et moléculaires) dont l'analyse pourra être intégrée au projet.

Rhésus			Gm		
Famille linguistique	Région	Nombre d'échantillons	Famille linguistique	Région	Nombre d'échantillons
SINO-TIBETAIN	Birmanie-Chine-Tibet-Taiwan	22	SINO-TIBETAIN	Bhoutan-Chine-Malaisie-Taiwan, Thaïlande	66
"ALTAIQUE"	Japon-Corée-Mandchourie-Sibérie	19	"ALTAIQUE"	Japon-Corée-Mandchourie-Mongolie	54
AUSTRO-NESIEN	Total	62	AUSTRO-NESIEN	Total	34
	Malaisie-Indonésie	6		Malaisie-Indonésie	6
	Taiwan	4		Taiwan	4
	Philippines	1		Philippines	4
	Micronésie	10		Micronésie	3
	Mélanésie	35		Mélanésie	15
	Polynésie	6		Polynésie	2
DAIQUE	Birmanie-Thaïlande	4	DAIQUE	Chine-Thaïlande	10
AUSTRO-ASIATIQUE	Asie du sud-est	5	AUSTRO-ASIATIQUE	Asie du sud-est	4
CAUCASIEN	Caucase	1	MIAO-YAO	Chine, Thaïlande	2
Total		113	Total		170
Moyenne des tailles d'échantillon		259.4	Moyenne des tailles d'échantillon		164.6

B.3.3. Démarche

La première phase d'obtention des données sera directement suivie d'une phase d'analyse.

Du point de vue linguistique, l'obtention des arbres datés de classification linguistique nous permettra de quantifier les relations de parenté entre langues, selon chacune des hypothèses linguistiques en présence, en termes de "distances" linguistiques relatives.

Du point de vue génétique, les données seront d'abord utilisées pour conduire une analyse globale des affinités génétiques des populations d'Asie orientale (analyses multivariées des distances génétiques, estimation d'indices de différenciation génétique).

Pour confronter les deux disciplines (linguistique et génétique), nous ferons appel à deux procédures:

1) corrélation entre processus de différenciation génétique et de différenciation linguistique des populations: cette procédure consiste à quantifier l'association observée entre affinités génétiques et linguistiques des populations, en calculant la corrélation entre distances génétiques et distances linguistiques. La procédure a été appliquée, par notre laboratoire et par d'autres, à l'échelle des continents Européen et Africain, ainsi qu'à l'échelle mondiale (Sokal 1988; Barbujani & Sokal 1990; Barbujani 1991; Excoffier et al. 1991; Sokal et al. 1992, 1996; Barbujani & Pilastro 1993; Chen et al. 1995; Weng & Sokal 1995; Poloni et al. 1997, 2000; Poloni 1999). Pour le projet de recherche, chaque hypothèse de regroupement linguistique de niveau supérieur donnera lieu à des estimations différentes des distances linguistiques entre paires de langues. Celles-ci seront utilisées pour calculer leur corrélation aux distances génétiques, afin d'estimer les coefficients de détermination des distances linguistiques sur les distances génétiques, indépendamment des distances géographiques (corrélations partielles). Nous espérons ainsi pouvoir favoriser l'une ou l'autre des hypothèses linguistiques en présence, sur la base des relations génétiques entre populations.

2) test de la perméabilité de frontières linguistiques au flux génique: cette procédure, développée et implémentée dans notre laboratoire par I. Dupanloup de Ceuninck (Dupanloup de Ceuninck 1999; Dupanloup de Ceuninck et al. in press), permet de quantifier la perméabilité (ou l'imperméabilité) de frontières linguistiques, et de différentes sections de ces frontières, à l'échange génétique entre populations. L'application de cette procédure permet d'affiner l'interprétation des corrélations entre génétique et linguistique, en testant si les frontières linguistiques correspondent à des barrières aux contacts génétiques entre populations. En effet, il est possible de supposer que la corrélation entre génétique et linguistique ait été induite par l'action des frontières linguistiques sur le flux de gènes entre populations, et non pas forcément (ou uniquement) par une synchronisation des processus de différenciation génétique et linguistique dans l'histoire des populations. D'autre part, la méthode permet également de localiser géographiquement des frontières génétiques, c'est-à-dire des zones où le taux de variation génétique est particulièrement important. Une fois localisées, ces frontières peuvent alors être comparées avec les frontières linguistiques déterminées par l'une ou l'autre des hypothèses linguistiques en présence.

Parce que la transmission de gènes d'une génération à la suivante est soumise à des processus stochastiques, chaque gène ou portion du génôme possède une histoire propre, plus ou moins indépendante. Pour inférer des scénarios relatifs à l'histoire du peuplement, il est donc nécessaire d'étudier plusieurs systèmes génétiques. Après avoir étudié les systèmes Gm et Rhésus, nous

nous proposons de répéter la même démarche de recherche à d'autres systèmes génétiques (sérologiques et moléculaires) que l'on jugera appropriés pour le projet (sur la base de critères portant sur la quantité et la qualité des données disponibles, ainsi que sur l'intérêt par rapport aux hypothèses de travail). Au vu de l'intérêt actuel pour l'étude du polymorphisme moléculaire humain (e.g. les projets "Human Genome Diversity"), de nouvelles données populationnelles s'accroissent à grande vitesse. Nous pensons donc pouvoir disposer très rapidement de données adéquates pour des polymorphismes moléculaires, tels que ceux de l'ADN mitochondrial et du chromosome Y.

Parallèlement, nous désirons aussi aborder la problématique des flux génétiques entre populations dans une perspective linguistique. En effet, par l'étude des emprunts entre langues, la linguistique permet de mettre en évidence des contacts, ponctuels ou de longue durée, entre populations (parlant des langues apparentées ou non), et ceux-ci peuvent parfois être datés, tout au moins sur une échelle relative. Une base de données sur les emprunts chinois dans les langues en contact est en cours de constitution au CRLAO dans le cadre d'une opération de recherche dirigée par L. Sagart et à laquelle collabore activement G. Jacques. Par ailleurs, le programme de test de la "perméabilité" des frontières linguistiques permet aussi la mise en évidence de contacts entre populations, attestés par les flux génétiques. Nous pensons donc qu'il sera intéressant de confronter les deux approches dans quelques zones d'intérêt, choisies d'après les résultats de la première phase de notre projet.

Enfin, nous évaluerons la nécessité d'une enquête de terrain dans population choisie en fonction du manque de données, de son intérêt pour le projet, et de nos contacts sur place (à Taiwan, en Chine, ou auprès des Kets de Sibérie?). Si l'enquête de terrain se révèle faisable et justifiée, nous évaluerons la possibilité d'échantillonner du matériel biologique (sang pour obtenir du serum, des cellules et de l'ADN, ou plus probablement frottis bucaux/cheveux pour obtenir de l'ADN) en fonction du budget alloué et du temps disponible.

B.4. Références bibliographiques

- Barbujani G. & Sokal, R.R. (1990) "Zones of sharp genetic change in Europe are also linguistic boundaries". *Proceedings of the National Academy of Sciences USA* 87: 1816-1819.
- Barbujani, G. (1991) "What do languages tell us about human microevolution?". *Trends in Ecology and Evolution* 6: 151-156.
- Barbujani, G. & Pilastro, A. (1993) "Genetic evidence on origin and dispersal of human populations speaking languages of the Nostratic macrofamily". *Proceedings of the National Academy of Sciences USA* 90: 4670-4673.
- Benedict, P.K. (1942) Thai, Kadai and Indonesian: a new alignment in Southeastern Asia. *American Anthropologist, n.s.*, 44, 576-601.
- Benedict, P.K. (1975) *Austro-Thai: language and culture*. HRAF Press. 490 p.
- Cavalli-Sforza, L.L., Menozzi, P. & Piazza A. (1994) *The History and Geography of Human Genes*. Princeton University Press, Princeton, New Jersey, pp 518.
- Chen, J., Sokal, R.R. & Ruhlen, M. (1995) "Worldwide analysis of genetic and linguistic relationships of human populations". *Human Biology* 67: 595-612.
- Chu, JY, Huang, W, Kuang, SQ, Wang, JM, Xu, JJ, Chu, ZT, Yang, ZQ, Lin, KQ, Li, P, Wu, M, Geng, ZC, Tan, CC, Du, RF, Jin L (1998) "Genetic relationship of populations in China". *Proceedings of the National Academy of Sciences USA* 95: 11763-11768.

- Dupanloup de Ceuninck, I. (1999) *Evaluation et synthèse des contributions de la linguistique et de la génétique à l'étude de la différenciation des populations humaines pendant la préhistoire récente*. Thèse de Doctorat: Laboratoire de Génétique et Biométrie, Université de Genève.
- Dupanloup de Ceuninck, I., Schneider, S., Langaney, A. & Excoffier, L. (in press) "Inferring the impact of linguistic boundaries on population differentiation: application to the Afro-Asiatic/Indo-European case". *European Journal of Human Genetics*.
- Excoffier, L., Harding, R.M., Sokal, R.R., Pellegrini, B. & Sanchez-Mazas, A. (1991) "Spatial differentiation of RH and GM haplotype frequencies in sub-Saharan Africa and its relation to linguistic affinities". *Human Biology* 63: 273-307.
- Hagelberg, E., Kayser, M., Nagy, M., Roewer, L., Zimdahl, H., Krawczak, M., Lio, P., Schiefenhover, W. (1999) "Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y chromosome and HLA markers". *Philos Trans R Soc Lond B* 354: 141-152.
- Hammer, MF & Horai, S (1995) " Y chromosomal DNA variation and the peopling of Japan". *American Journal of Human Genetics* 56: 951-962.
- Hashimoto, M. J. (1986) The altaicization of northern Chinese. in: John McCoy and Timothy Light (eds.) *Contributions to Sino-Tibetan Studies*, 76-97. Leiden: E. J. Brill.
- Karafet, T, Zegura, SL, Vuturo-Brady, J, Posukh, O, Osipova, L, Wiebe, V, Romero, F, Long, JC, Harihara, S, Jin, F, Dashnyam, B, Gerelsaikhan, T, Omoto, K, Hammer, MF (1997) " Y chromosome markers and Trans-Bering Strait dispersals ". *American Journal of Physical Anthropology* 102: 301-314.
- Karafet, TM, Zegura, SL, Posukh, O, Osipova, L, Bergen, A, Long, J, Goldman, D, Klitz, W, Harihara, S, de Knijff, P, Wiebe, V, Griffiths, RC, Templeton, AR, Hammer, MF (1999) " Ancestral Asian source(s) of new world Y-chromosome founder haplotypes". *American Journal of Human Genetics* 64: 817-831.
- Kim, W, Shin, DJ, Harihara, S, Kim, YJ (2000) " Y chromosomal DNA variation in east Asian populations and its potential for inferring the peopling of Korea". *Journal of Human Genetics* 45: 76-83.
- Matsumoto, H. (1988a) "Characteristics of Mongoloid and neighboring populations based on the genetic markers of human immunoglobulin ". *Human Genetics* 80:207-218.
- Matsumoto, H. (1988b) "Characteristics of Mongoloid populations based on the human immunoglobulin allotypes". *Anthrop. Anz* 46:119-127.
- Melton T, Clifford S, Martinson J, Batzer M, Stoneking M (1998) " Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes", *American Journal of Human Genetics* 63: 1807-1823.
- Poloni, E.S., Semino, O., Passarino, G., Santachiara-Benerecetti, A.S., Dupanloup, I., Langaney, A. & Excoffier, L. (1997) "Human genetic affinities for Y-chromosome p49a,f/TaqI haplotypes show strong correspondence with linguistics". *American Journal of Human Genetics* 61: 1015-1035.
- Poloni, E.S. (1999) *Polymorphismes de l'ADN et histoire du peuplement humain: apports de l'étude des marqueurs RFLP*. Thèse de Doctorat: Laboratoire de Génétique et Biométrie, Université de Genève.
- Poloni E.S., Ray, N., Schneider, S. & Langaney, A (2000) "Languages and genes: modes of transmission observed through the analysis of male-specific and female-specific genes". *The Evolution of Language, 3rd International Conference*, Paris: Ecole Nationale Supérieure des Télécommunications, 3-6 April 2000, abstract.
- Pulleyblank, E. (1995) The historical and prehistorical relationships of Chinese. In: William S.-Y. Wang (ed.) *The Ancestry of the Chinese Language*. Journal of Chinese Linguistics monograph series no. 8, pp. 145-194.
- Reid, L. (1994) Morphological evidence for Austric. *Oceanic Linguistics* 33, 2: 323-344.
- Sagart, L. (1994) "Old Chinese and Proto-Austronesian evidence for Sino-Austronesian". *Oceanic Linguistics* 33 (2): 271-308.
- Sanchez-Mazas, A. (1990) *Polymorphisme des systèmes immunologiques Rhésus, GM et HLA et histoire du peuplement humain*. Thèse de Doctorat: Laboratoire de Génétique et Biométrie, Université de Genève.
- Sanchez-Mazas, A. & Pellegrini, B. (1990) "Polymorphismes Rhésus, Gm et HLA et histoire de l'Homme moderne". *Bull. et Mém. de la Soc. d'Anthrop. de Paris, n.s.*, 2 (1): 57-76.
- Schmidt, W. (1906) *Die Mon-Khmer Voelker, ein Bindeglied Zwischen Voelkern Zentralasiens und Austronesiens*. Braunschweig: Friedrich Vieweg und Sohn.
- Sokal, R.R. (1988) "Genetic, geographic, and linguistic distances in Europe". *Proceedings of the National Academy of Sciences USA* 85: 1722-1726.
- Sokal, R.R., Oden, N.L. & Thomson, B.A. (1992) "Origins of the Indo-Europeans: genetic evidence". *Proceedings of the National Academy of Sciences USA* 89: 7669-7673.

- Sokal, R.R., Oden, N.L., Walker, J., Di Giovanni, D., Thomson, B.A. (1996) "Historical population movements in Europe influence genetic relationships in modern samples". *Human Biology* 68: 873-898.
- Starostin, S. A. (1984 [1987]) *A hypothesis on genetic links of the Sino-Tibetan languages with the Yeniseian and north-Caucasian languages*. Translated and annotated by W. H. Baxter III. Paper circulated at the 20th International Conference on Sino-Tibetan Languages and Linguistics, Vancouver, 20-23 August 1987.
- Su, B, Xiao, J, Underhill, P, Deka, R, Zhang, W, Akey, J, Huang, W, Shen, D, Lu, D, Luo, J, Chu, J, Tan, J, Shen, P, Davis, R, Cavalli-Sforza, L, Chakraborty, R, Xiong, M, Du, R, Oefner, P, Chen, Z, Jin, L (1999) " Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age". *American Journal of Human Genetics* 65: 1718-1724.
- van Driem, G. (1995) Black Mountain conjugational morphology, Proto-Tibeto-Burman morphosyntax, and the linguistic position of Chinese. In: Yoshio Nishi, James A. Matisoff and Yasuhiko Nagano (eds.) *New Horizons in Tibeto-Burman morphosyntax*. Senri ethnological studies 41: 229-259. Osaka: National Museum of Ethnology.
- Vovin, A. (1994) "Is Japanese related to Austronesian ?" *Oceanic Linguistics* 33, 2: 269-390.
- Weng, Z. & Sokal, R.R. (1995) "Origins of Indo-Europeans and the spread of agriculture in Europe: comparison of lexicostatistical and genetic evidence". *Human Biology* 67: 577-594.
- Zerjal, T, Dashnyam, B, Pandya, A, Kayser, M, Roewer, L, Santos, FR, Schiefenhover, W, Fretwell, N, Jobling, MA, Harihara, S, Shimizu, K, Semjidmaa, D, Sajantila, A, Salo, P, Crawford, MH, Ginter, EK, Evgrafov, OV, Tyler-Smith, C (1997) " Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis". *American Journal of Human Genetics* 60: 1174-1183.
- Zhao, T. & Lee, T.D. (1989) "Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation". *Human Genetics* 83: 101-110.